ACL 2023

The Fourth Workshop on Insights from Negative Results in NLP

Proceedings of the Workshop

May 5, 2023

The ACL organizers gratefully acknowledge the support from the following sponsors.

Silver

Google

©2023 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-959429-49-4

Introduction

Publication of negative results is difficult in most fields, and the current focus on benchmark-driven performance improvement exacerbates this situation and implicitly discourages hypothesis-driven research. As a result, the development of NLP models often devolves into a product of tinkering and tweaking, rather than science. Furthermore, it increases the time, effort, and carbon emissions spent on developing and tuning models, as the researchers have little opportunity to learn from what has already been tried and failed.

Historically, this tendency is hard to combat. ACL 2010 invited negative results as a special type of research paper submissions[1], but received too few submissions and did not continue with it. *The Journal for Interesting Negative Results in NLP and ML*[2] has only produced one issue in 2008.

However, the tide may be turning. The fourth iteration of the *Workshop on Insights from Negative Results* attracted 25 submissions and 10 from EACL 2023 Findings.

The workshop maintained roughly the same focus, welcoming many kinds of negative results with the hope that they could yield useful insights and provide a much-needed reality check on the successes of deep learning models in NLP. In particular, we solicited the following types of contributions:

- broadly applicable recommendations for training/fine-tuning, especially if X that didn't work is something that many practitioners would think reasonable to try, and if the demonstration of X's failure is accompanied by some explanation/hypothesis;
- ablation studies of components in previously proposed models, showing that their contributions are different from what was initially reported;
- datasets or probing tasks showing that previous approaches do not generalize to other domains or language phenomena;
- trivial baselines that work suspiciously well for a given task/dataset;
- cross-lingual studies showing that a technique X is only successful for a certain language or language family;
- experiments on (in)stability of the previously published results due to hardware, random initializations, preprocessing pipeline components, etc;
- theoretical arguments and/or proofs for why X should not be expected to work;
- demonstration of issues with data processing/collection/annotation pipelines, especially if they are widely used;
- demonstration of issues with evaluation metrics (e.g. accuracy, F1 or BLEU), which prevent their usage for fair comparison of methods.

In terms of topics/themes, 6 papers from our accepted proceedings discussed "Representation Learning / Pre-training"; 1 discussed "Entity Detection/Resolution"; 1 paper examined text classification; 1 dealt with issues of robustness, generalizability, error analysis; 2 on the topic of "text comprehension / VQA"; 2 papers focused on text generation such as summarization, machine translation; 1 on replication of human evaluations in NLP. Some submissions fit in more than one category.

We accepted 14 short papers (56.0% acceptance rate) and 10 papers from EACL 2023 Findings.

We hope the workshop will continue to contribute to the many reality-check discussions on progress in NLP. If we do not talk about things that do not work, it is harder to see what the biggest problems are and where the community effort is the most needed.

[1] https://mirror.aclweb.org/acl2010/papers.html

[2] http://jinr.site.uottawa.ca/

Organizing Committee

Organizers

Shabnam Tafreshi, University of Maryland: ARLIS, USA
Arjun Reddy Akula, Google, USA
João Sedoc, New York University, USA
Anna Rogers, University of Copenhagen, Denmark
Aleksandr Drozd, RIKEN, Japan
Anna Rumshisky, University of Massachusetts Lowell / Amazon Alexa, USA

Program Committee

Chairs

Arjun Akula, Google, USA
Aleksandr Drozd, RIKEN Center for Computational Science
Anna Rogers, University of Copenhagen
Anna Rumshisky, University of Massachusetts Lowell
João Sedoc, New York University
Shabnam Tafreshi, UMD:ARLIS

Program Committee

Amittai Axelrod, Apple
Ameya Godbole, University of Southern California
Andrey Kutuzov, University of Oslo
Anil Nelakanti, Amazon
Ali Seyfi, The George Washington University
Anuj Khare, Google LLC
Arijit Adhikari, Amazon
Ashutosh Modi, Indian Institute of Technology Kanpur
Chanjun Park, Upstage
Chung-chi Chen, National Institute of Advanced Industrial Science and Technology
Constantine Lignos, Brandeis University
David Samuel, University of Oslo, Language Technology Group
Edison Marrese-taylor, National Institute of Advanced Industrial Science and Technology (AIST)
Efsun Sarioglu Kayi, Johns Hopkins APL
Emil Vatai, Riken R-CCS
Gaurav Mishra, Google
Giovanni Puccetti, Scuola Normale Superiore di Pisa
Guenter Neumann, DFKI
Saarland University
Happy Buzaaba, RIKEN
John Ortega, Northeastern University
Joinal Ahmed, Google
Kaveri Anuranjana, Saarland University
Mahesh Goud Tandarpally, Amazon
Marzena Karpinska, University of Massachusetts Amherst
Maximilian Spliethöver, Leibniz University Hannover
Neha Nayak Kennard, University of Massachusetts Amherst
Salvatore Giorgi, University of Pennsylvania
Shubham Chatterjee, University of Glasgow
Tamás Ficsor, University of Szeged
Tristan Naumann, Microsoft Research
Wazir Ali, Institute of Business Management

Keynote Talk: My fruitless endeavours with neuro-symbolic NLP

Vered Shwartz
University of British Columbia, Canada

Bio: Vered Shwartz is an Assistant Professor of Computer Science at the University of British Columbia, and a CIFAR AI Chair at the Vector Institute. Her research interests include commonsense reasoning, computational semantics and pragmatics, and multiword expressions. Previously, Vered was a postdoctoral researcher at the Allen Institute for AI (AI2) and the University of Washington, and received her PhD in Computer Science from Bar-Ilan University. Vered's work has been recognized with several awards, including The Eric and Wendy Schmidt Postdoctoral Award for Women in Mathematical and Computing Sciences, the Clore Foundation Scholarship, and an ACL 2016 outstanding paper award.

Keynote Talk: Do not give up on projects with negative results!

Mohit Iyyer
UMass Amherst, USA

Bio: Mohit Iyyer is an assistant professor in computer science at the University of Massachusetts Amherst. His research focuses broadly on designing machine learning models for discourse-level language generation (e.g., for story generation and machine translation), and his group also works on tasks involving creative language understanding (e.g., modeling fictional narratives and characters). He is the recipient of best paper awards at NAACL (2016, 2018) and a best demo award at NeurIPS 2015, and he received the 2022 Samsung AI Researcher of the Year award. He received his PhD in computer science from the University of Maryland, College Park in 2017, advised by Jordan Boyd-Graber and Hal Daumé III, and spent the following year as a researcher at the Allen Institute for Artificial Intelligence.

Keynote Talk: How negative results fuel our research: insights from gender bias and multilinguality

Hila Gonen
University of Washington, USA

Bio: Hila is a postdoctoral Researcher at Meta AI and at the Paul G. Allen School of Computer Science Engineering at the University of Washington. Hila's research lies in the intersection of Natural Language Processing, Machine Learning and AI. She is interested in analyzing and better understanding the cutting-edge technology used in the field, and focuses mainly on multilinguality and fairness in her research. Before joining UW and Meta AI, Hila was a postdoctoral researcher at Amazon. Prior to that she did her Ph.D in Computer Science at the NLP lab at Bar Ilan University. She obtained her Ms.C. in Computer Science from the Hebrew University. Hila is the recipient of several postdoc awards and an EECS rising stars award. Her work received the best paper awards at CoNLL 2019 and at the repL4nlp workshop 2022, and also an outstanding thesis award from IAAI.

Keynote Talk: Three lessons from negative results in NLP research

Rachel Rudinger
University of Maryland, USA

Bio: Rachel is an assistant professor at university of Maryland. Her research is focused on problems in natural language understanding, including knowledge acquisition, commonsense reasoning, an semantic representation.

Table of Contents

Program

Friday, May 5, 2023

09:00 - 09:15 *Opening Remarks*

09:15 - 10:00 *Thematic Session 1: Text Generation*

Hiding in Plain Sight: Insights into Abstractive Text Summarization
Vivek Srivastava, Savita Bhat and Niranjan Pedanekar

Encoding Sentence Position in Context-Aware Neural Machine Translation with Concatenation
Lorenzo Lupo, Marco Dinarelli and Laurent Besacier

Exploring the Reasons for Non-generalizability of KBQA systems
Sopan Khosla, Ritam Dutt, Vinayshekhar Bannihatti Kumar and Rashmi Gangadharaiah

10:00 - 10:45 *Invited Talk: Vered Shwartz*

10:45 - 11:15 *Coffee Break*

11:15 - 11:45 *Thematic Session 2: Text Classification & Comprehension*

Can Demographic Factors Improve Text Classification? Revisiting Demographic Adaptation in the Age of Transformers
Chia-Chien Hung, Anne Lauscher, Dirk Hovy, Simone Ponzetto and Goran Glavas

A Data-centric Framework for Improving Domain-specific Machine Reading Comprehension Datasets
Iva Bojic, Josef Halim, Verena Suharman, Sreeja Tar, Qi Chwen Ong, Duy Phung, Mathieu Ravaut, Shafiq Joty and Josip Car

An Empirical Study on Active Learning for Multi-label Text Classification
Mengqi Wang and Ming Liu

11:45 - 12:15 *Thematic Session 3: Representation Learning & Pre-training*

SocBERT: A Pretrained Model for Social Media Text
Yuting Guo and Abeed Sarker

Friday, May 5, 2023 (continued)

Edit Aware Representation Learning via Levenshtein Prediction
Edison Marrese-taylor, Machel Reid and Alfredo Solano

What changes when you randomly choose BPE merge operations? Not much.
Jonne Saleva and Constantine Lignos

12:15 - 14:00 *Lunch*

14:00 - 14:30 *Invited Talk: Mohit Iyyer*

14:30 - 15:00 *Thematic Session 4: Robustness & Error Analysis*

Benchmarking Long-tail Generalization with Likelihood Splits
Ameya Godbole and Robin Jia

Transformer-based Models for Long-Form Document Matching - Challenges and Empirical Analysis
Akshita Jha, Adithya Samavedhi, Vineeth Rakesh, Jaideep Chandrashekar and Chandan Reddy

Annotating PubMed Abstracts with MeSH Headings using Graph Neural Network
Faizan Mustafa, Rafika Boutalbi and Anastasiia Iurshina

15:00 - 15:30 *Invited Talk: Rachel Rudinger*

15:30 - 16:00 *Coffee Break*

16:00 - 16:30 *Invited Talk: Hila Gonen*

16:30 - 18:00 *Poster Session*

18:00 - 18:10 *Closing Remarks*